INTERMODULATION DISTORTION: A CONTROLLABLE PARAMETER IN THE ANALYSIS OF THE INTELLIGIBILITY OF CLIPPED SPEECH

WILLIAM JOSEPH LANNES

INTERMODULATION DISTORTION:

A CONTROLLABLE PARAMETER IN THE ANALYSIS

OF THE INTELLIGIBILITY OF CLIPPED SPEECH

by

William Joseph Lannes, III
Captain, United States Marine Corps
B.S.E.E., Tulane University, 1959

Submitted in partial fulfillment
for the degree of

MASTER OF SCIENCE IN ELECTRONICS ENGINEERING

from the

UNITED STATES NAVAL POSTGRADUATE SCHOOL
May 1966

ABSTRACT

Until the experiments conducted on Single Sideband speech clipping in the early 1960's, very little work had been done in the field of conventional speech processing or clipping since the late 1940's. Although the SSB experiments were mainly motivated by the unique repeaking problem associated with clipped SSB signals, they pointed out that further improvements in intelligibility of clipped speech were still possible. Perhaps of more importance, these experiments offered experimental evidence that the problem of intelligibility of clipped speech was closely related to the intermodulation distortion produced in the clipping process. This paper is a study of clipped speech as viewed from this point. The relation of the intermodulation distortion to the intelligibility of clipped speech is developed, and it is shown that intelligibility can be enhanced by reduction of the intermodulation products. A mathematical approach is developed, resulting in a simple

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Shortly after World War II there was a great deal of interest in improving the capabilities of our communications systems. Much testing and study was done in the area of speech processing. Although some of these speech processing circuits involved integraters, differentiators, and filtering techniques, the circuits which received the most attention were what might be considered as conventional audio clippers. These were simply diodes with external voltage sources to establish the clipping levels. A great amount of useful information about speech clipping was amassed during this period. Whereas this information was invaluable to the design engineer, it was empirical in nature and (with the exception of a few studies) it supplied very little of the "whys" concerning the tabulated results.

During the 1950's interest in speech clipping appeared to wane; and indeed, it seemed as if most of the work to done in this area had been completed. Study groups and individuals interested in speech processing in general began shifting their efforts to such promising techniques as the vocoder systems.

It was about this time, however, that interest was renewed in Single Sideband as a mode of communications. This was due in part, to the availability of good mechanical filters and small stable oscillators. With SSB, came a renewed interest in speech clipping. The audio clipping techniques perfected in the late 1940's worked admirably in AM systems, but caused severe repeaking in SSB systems. Experiments in clipping the SSB signal instead of the audio signal resulted in improvements in the repeaking problem. Additionally, it found that clipping the SSB signal gave higher intelligibility than the clipped audio signal. This is an interesting result since the same clipping levels and same equipments were used in the audio clipping and SSB clipping comparisons. It seems logical, therefore, that it is the elimination of the intermodulation products which fall outside of the pass band in SSB clipping that causes the improvement in intelligibility. In audio clipping most of these products fall in the pass band. The following study will show, however, that this technique of improving the intelligibility of clipped speech by clipping at R-F rather than audio will only give good results for SSB R-F signals and not DSB signals.

The following three chapters of this investigation are related in that they all deal with the considerations involved in discussing the intelligibility of clipped speech as a function of the intermodulation products. The results of the post WW II tests, the implications of the SSB tests, and certain aspects of the theories of speech and hearing are combined

in order to clearly define the problem of intelligibility of clipped speech. Having accomplished this, the remaining two chapters deal with presenting ways of solving this problem. That is, specific means to improve the intelligibility of clipped speech are presented.

CHAPTER II

DISTORTION IN CLIPPED SPEECH

All nonlinear systems produce distortion of some type. In the clipping process this distortion is so obvious that it is surprising that the clipped output yields any of the information that was contained in the original waveform. A close examination of a speech waveform that has been subjected to severe symmetrical clipping will indicate that the waveform is essentially binary in nature. That is, it consists of positive and negative pulses of amplitude equal to the clipping level. This waveform is drastically different from the normal (unclipped) speech waveform.

There are three types of distortion that are possible when a signal is processed. These are amplitude, frequency, and phase distortion. As might be expected all three are present in the clipping process. The job at hand, then, is to determine how much, if at all, each type of distortion affects intelligibility.

Perhaps before we look closer at the distortion in clipping it might be well to say something about intelligibility. Intelligibility as it is referred to in this paper is a measure of how much of the original information is assimilated by the listner. Intelligibility, therefore, implies

only whether or not the spoken message is understood and says nothing about the "naturalness" of the speech. For example, clipping followed by filtering can noticeably affect the quality of speech sounds without altering any of the message content.

The most obvious type of distortion present in clipping is amplitude distortion. This is hardly unexpected since the clipping operation is essentially an amplitude limiting process. Additionally, most practical circuits include filters following the clipper (to maintain the original bandwidth) and therefore the amplitude distortion introduced by these filters should also be considered. The amplitude distortion due to the filter is caused both by the phase characteristics of the filter and by the rejection by the filter of the higher order frequency components that are present in clipped waves. Nonlinear phase characteristics cause the components of the filtered wave to add (vectorally) in a slightly different manner than the original components. This results in different amplitudes even for waveforms which have the same frequency components. Although, compared to the clipping levels, the distortion due to the filter is small. However, it can not be neglected. Filtering following the clipper will generally raise the peak to average power ratio by 3DB.[1]

[1]E. W. Pappenfus, W. B. Bruene, and E. O. Schoenike, Single Sideband Principles and Circuits (New York: McGraw-Hill Book Company, 1964), p. 328.

This is due mostly to the rejection of the higher order frequencies. Peaking or flattening due to the phase characteristics only is generally of the order of 1 DB for a typical SSB filter.[2]

Because of the amplitude limiting nature of the clipper it seems reasonable to assume that it is this severe amplitude distortion which has the greatest effect on intelligibility. Appealing as this assumption is to our intuition, tests indicate that amplitude distortion is not a primary factor in intelligibility of clipped speech. Many impressive experiments have been conducted to emphasize this fact. One such experiment consists of a circuit which puts out only positive and negative pulses (equal amplitude) and is synchronized to the zero crossings of the original speech wave. The synthetic wave produced (using only the frequency information of the original waveform) is essentially that of a severely clipped speech waveform. Despite the fact that the quality is poor, this synthetic waveform yields intelligible speech.[3] This indicates (this also has been demonstrated in different ways by others) that even speech that has had its envelope so severely clipped

[2]Air Force Cambridge Research Laboratories, Speech-Signal Processing and Applications to Single Sideband (Bozeman: AD-276 850, Montana State College, 1962), p. 99.

[3]J. C. Licklider, and I. Pollack, "Effects of Differentiation, Integration, and Infinite Peak Clipping upon the Intelligibility of Speech," Journal of Acoustical Society of America, Vol. 20, No. 3, 1948, p. 42.

that only the zero crossings are retained still remains intelligible.[4]

To state that the amplitude distortion introduced by the clipping process has little effect on intelligibility would really only be correct for a noiseless system and even then it would depend somewhat on the level of clipping. When we consider a system with noise, which indeed we must, what we really mean is that the severe distortion of the shape of the waveform due to the amplitude limiting process has little effect on intelligibility. That is, even in the presense of noise, we can restore the amplitude (not shape) of the clipped wave; transmit it and then receive it as intelligible speech. In fact, if it is a particularly noisy system, we will probably increase the intelligibility by clipping the signal prior to amplifying it for transmission. This result indicates that while the energy or power in the wave (which is a function of amplitude) and its frequency (zero crossings) are important to intelligibility the shape of the waveform is not. This also indicates the importance of the signal to noise ratio to the level of intelligibility. In summary then we can conclude that amplitude distortion, that is distortion of the waveshape, has neglible effects on intelligibility.

As was previously mentioned, the filters in the clipping systems introduce phase distortion. Although this phase distortion is small it

[4]J. M. C. Dukes, "The Effect of Severe Amplitude Limitation on Certain Types of Random Signal: A Clue to the Intelligibility of 'Infinitely' Clipped Speech," Proceedings of the IEE (London), Vol. 102-103, Pt. C, p. 88.

must be considered. Many experiments have been conducted specifically to study how the human ear responds to phase difference.[5] These tests indicate that although the phase difference can be preceived by the human ear, it can be done only in a qualitative sense. That is, tones that are out of phase have a slightly different quality which is preceiveable by the human ear. Since this is a qualitative effect it does not affect the intelligibility; this is probably what Flecther means when he states that the ear does not ordinarily recognize phase difference.[6]

To illustrate this point, oscillograms have been recorded of speech waveforms which have been subjected to severe phase distortion. When compared to the shape of the original waveform there is generally little agreement; yet both waveforms can be recognized as the same sound.[7] Thus, as was implied earlier, the phase distortion in speech waveforms might be considered as amplitude distortion since it results in a change of shape of the waveform. It is interesting to note that later we will find that intelligibility depends almost entirely on the power spectrum of the speech waveform. The power spectrum gives no information about the phase of the frequency components and therefore agrees with the conclusion that the phase has little effect on intelligibility.

[5]E. G. Wever, Theory of Hearing (New York: John Wiley & Sons, Inc., 1949), p. 425.

[6]Harvey Fletcher, Speech and Hearing in Communications (New York: D. Van Nostrand Company, 1953), p. 32.

[7]Ibid., p. 48.

Now that it has been established that amplitude and phase distortion are not the primary causes of loss of intelligibility we must examine the one remaining type of distortion, frequency distortion. Frequency distortion is introduced in the clipping process in the form of inter-modulation (IM) products. The nature of the IM products is such that in addition to the original, desired frequencies, new or undesired frequencies are generated as harmonics and sums and differences of the original signals and their harmonics.[8] These IM products are present in all nonlinear systems where the input signal has at least two frequency components. Since speech waveforms consist of many frequency components and since the clipping characteristic represents a severe non-linearity, the IM phenomenia is pronounced.

Having eliminated two of the three possible types of distortion as the probable sources of difficulty, it might reasonably be assumed that the IM phenomenia is the source of trouble. However, since this is of such importance to the study of clipped speech we want to take a much closer look as to why this is true. Before we embark on a discussion of how IM products affect the intelligibility of clipped speech we should first examine some of the factors governing the intelligibility of normal (unprocessed) speech.

[8]See Chapter 5.

CHAPTER III

THE NATURE OF SPEECH AND HEARING

Speech

Speech may be broadly classified as voiced and unvoiced sounds. The voiced sounds are the vowels and the unvoiced sounds are the consonants. Between the vowels and consonants are such classes of sounds as the dipthongs and semi-vowels. These are transitional sounds and they contain properties that are sometimes common to both vowels and consonants. Since the vowels and consonants are the most important of the speech sounds only they will be considered in the present discussion.

The vowels or voiced sounds are produced by the vibration of the vocal chords. These voiced sounds consist primarily of harmonics of the fundamental frequency of the vocal chords. The fundamental frequency is approximately 70 to 250 cps for males, and as high as 350 cps for females.[1] In the formation of the different vowel sounds the vocal cavities play an important role. These cavities are the oral,

[1]A. J. Strassman, and K. C. Stockhoff, "Military Applications for Speech Compression Techniques," Hughes Aircraft Company, OP-24, April, 1960, p. 2.

throat or pharynx, and nasal cavities. The oral cavity being the most important in that it can be controlled between wide limits by movement of the tongue and lower jaw. These cavities act as resonators and as such reinforce certain harmonics of the fundamental frequency. This results in two and sometimes three (seldom one) frequency regions of prominence in the speech spectrum. The relative location of these regions of accentuated intensity, known as formants, is an important characteristic in the difference between vowel sounds.[2] It is generally necessary to retain these distinctive harmonic structures or the character of many sounds will be altered or lost.[3] It is important to realize that the formant frequencies do not occur at exactly the same frequencies when the same vowel is spoken by different persons. Work by Potter and Steinberg of Bell Laboratories has indicated that all the formant frequencies are slightly higher for a woman's voice and considerably higher for a child's voice. Thus it appears that it is the relation between the harmonic frequencies of each formant region and the manner in which their intensities are distributed (i.e., shape) which are characteristic of the vowels rather than the absolute values of the formant frequencies. This is supported by the fact that the

[2]Harvey Fletcher, Speech and Hearing in Communications (New York: D. Van Nostrand Company, 1953), p. 10.

[3]E. W. Pappenfus, W. B. Bruene, and E. O. Schoenike, Single Sideband Principles and Circuits (New York: McGraw Hill Book Company, 1964), p. 324.

manner of starting and stopping a vowel gives more information that may be used to recognize the vowel than is given by the absolute values of the formant frequencies.[4] Since the shape of the frequency spectrum is so important to intelligibility it might be well to say something about the spectrum itself. For signal analysis it is necessary to define the spectral densities slightly differently for different types of signals (i.e., periodic, pulse, random, etc.). It will, however, suffice for the purpose of this discussion to state that the spectrum prepresents a plot of the relative intensities at the signal's component frequencies. In other words, it represents measureable parameters such as might be obtained by using a spectrum or wave analyzer.

The formant regions are caused by the vowel sounds and since they represent high intensity regions it is not suprising to find that the vowels constitute the major portion of intelligible speech. Let us see what effect the consonants have.

The consonants or unvoiced sounds are formed by forcing an air stream through narrow orifices and past sharp edges. Different positions of the tongue, teeth, and lip provide the necessary passages. The formation of consonants does not involve the vibration of the vocal chords, consequently consonants do not have regions of reinforced intensity (formant regions) characteristic of the vowel sounds. Although

[4]Harvey Fletcher, op. cit., p. 61.

when words consist of combinations of vowels and consonants there is a subtle relation between the presence of the consonants and the shape of the vowel produced formant regions. Consonants also have no harmonic relation existing between the frequency components. In fact, consonants are generally considered as having a random frequency distribution. Although consonants are attributed with this random property the intensity or power of the consonants is concentrated in the high frequency region.

The intensity of the consonants are of interest because intensity is one of the two most important characteristics of speech: the other characteristic being frequency. To remain consistent with the earlier discussion of spectral densities the intensity of the wave form may be thought of as synonomous with either energy or power. Both are direct functions of intensity; however, for a comparison of intensities the units must be the same such as DB difference between components when all are measured in watts. The importance of intensity was inferred earlier when it was stated that the vowels constituted the major portion of intelligible speech due to their inherent high intensity. Physically this can be interperted as meaning that the lungs are acting in their full capacity in establishing the vibrating columns of air characteristic of the vowels. This is not the case for consonants. The air stream previously mentioned for producing the unvoiced sounds does

not always come directly from the lungs. Frequently only the air present in the mouth cavity is utilized in producing the consonants. If we consider the lungs to be the power supply for the physical system, it is easy to see why the consonants are low power or low intensity sounds and that they are easily masked by system noise.

Let us now consider the preceeding observations about the relative intensities of vowels and consonants in a normal communication system. By normal it is meant that the transmitter performs no nonlinear or non-uniform processes on the signal such as clipping or preemphasis and that the first stage of the receiver has a finite amount of noise (i.e. system is not noiseless). In such a system the vowels have a high probability of reception whereas it is very likely that most or perhaps even all of the consonants will be masked. This suggests that we can do nearly as well by only transmitting the vowels. Further, since the vowels are characterized by the formant regions it appears that if we could transmit only the information in these regions we could still get intelligible speech. This is, in fact, the basis of a special type of speech processer known as a formant vocoder.

The speech processing that is performed in the vocoder systems leads to reduced bandwidth of the transmitted signal whereas in clipping speech we are striving to reduce the peak to average ratio of the speech without noticeably affecting intelligibility. The vocoder system is of

interest, however, to our inquiry of the nature of speech. The main reason is that it offers experimental evidence of the main thought that we would like to take out of our discussion of speech; that is, it offers experimental evidence that the information is speech is carried largely in the varying shape of the power density spectrum and not in the sound pressure vs time plot normally seen on an oscilloscope.[5] The vocoder systems are of secondary interest in that they represent an electrical analogue to the human speech system. This analogue may help us to better understand the nature of speech itself.

The vocoder was first introduced by Homer Dudley in 1939. Dudley's vocoder was actually a synthetic speaker. By continually adjusting the controls, a skilled operator could make the machine "speak". A schematic diagram of an early vocoder and its human counterpart is shown in Figure 1.

The components of the vocoder are labeled in the figure and the relation to the human counterparts are indicated. The buzzer or "buzz generator" provides periodic sounds whose fundamental frequency can be varied and which corresponds to the pitch of the voice (fundamental frequency of the vocal chords). The randon noise source or "hiss generator" corresponds to the sounds made by consonants.

In spite of the fact that Dudley's system produced intelligible

[5]A. J. Strassman, and K. C. Stockhoff, op. cit., p. 1.

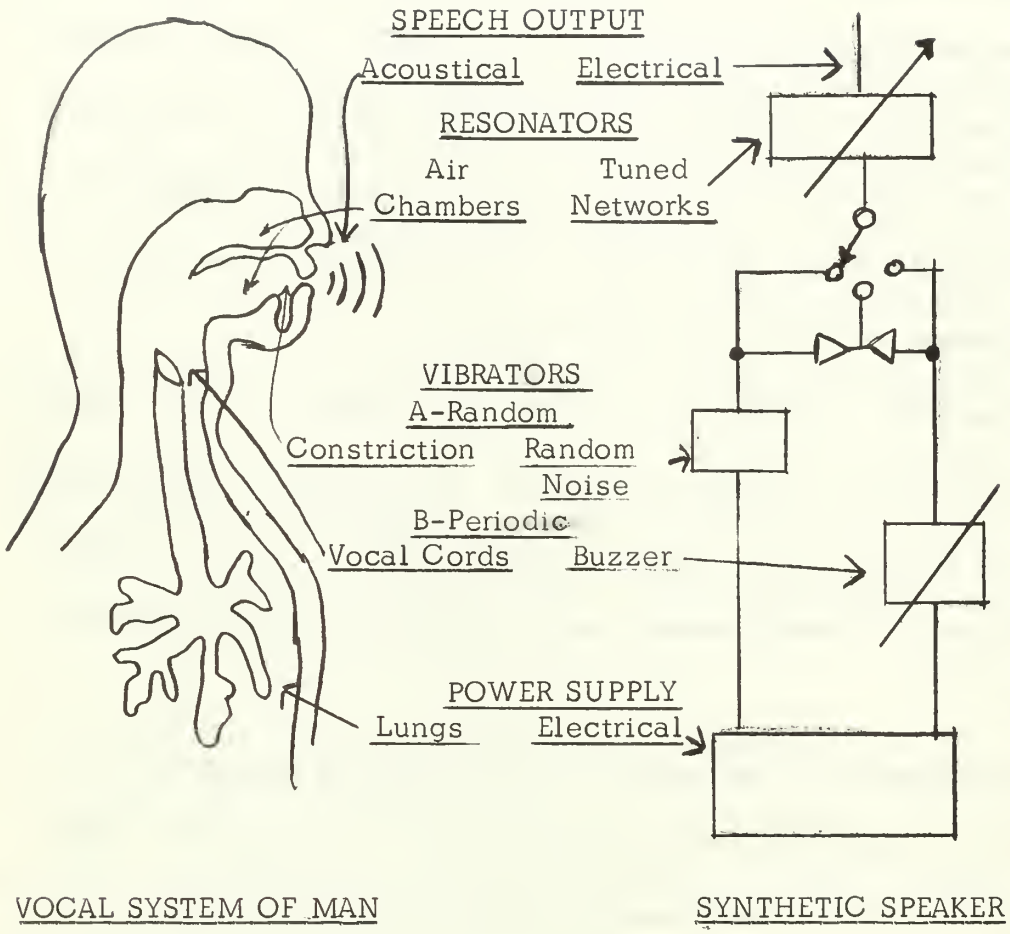


Figure 1. Comparison of Human Vocal System with Synthetic Speaker

speech it was of little practical value until high speed circuitry became reliable. The vocoder is now incorporated into complete speech processing systems for voice communication.

There are two common types of communication vocoders; the channel vocoder and the formant vocoder. In the channel vocoder system the speech bandwidth to be transmitted (for example 300 to 3500 cps) is divided into several sections (usually six or more) by selective filters, The energy distribution of each section is determined by spectrum analyzers. This signal (information from the spectral density) is then transmitted. In other words only the information as to frequency and intensity is transmitted to the receiving station. Since the maximum rate of change of energy in each section is related to the syllabic rate (about 25 cps), the dynamic spectrum information can be transmitted in a much narrower bandwidth than the original speech.[6] On the receiving end this signal is synthesized to recreate the original speech signal. The voice synthesizing portion of the vocoder operates on the same principle as Dudley's vocoder.

The formant vocoder operates on the same principle as the channel type except that the speech is approximated by utilizing only the information in the formant regions. By only transmitting the regions corresponding to the spectral peaks the bandwidth can be even further reduced.

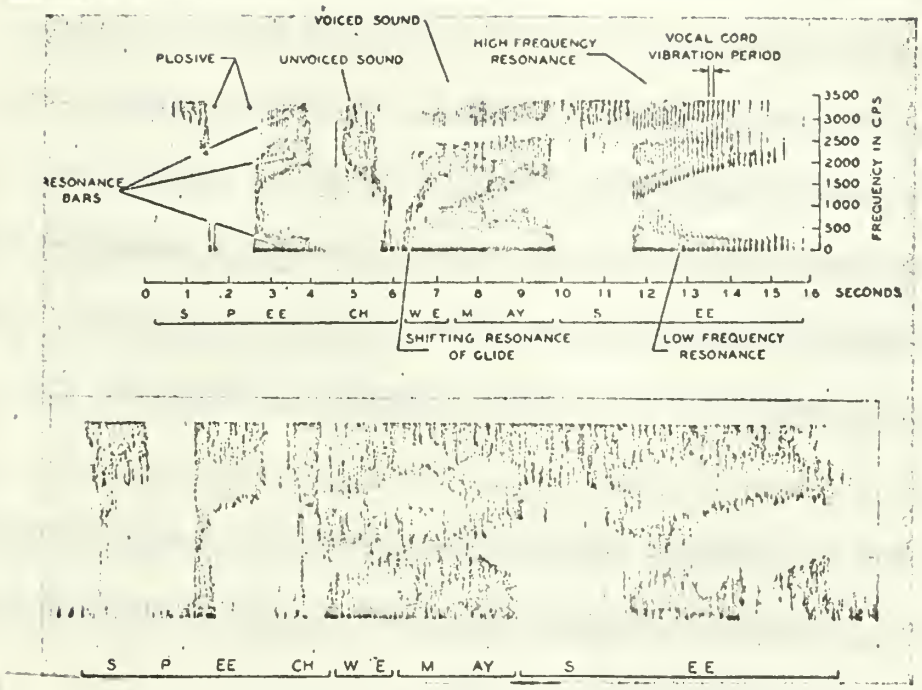
[6]E. W. Pappenfus, op. cit., p. 337.

The intelligibility of speech from formant vocoders in somewhat poorer that that from the channel vocoder. This is due mostly to the difficulty in accurately locating the formant frequencies.[7] As was previously mentioned these regions vary with different speakers. The fact, however, that the formant vocoder produces intelligible speech is additional proof that most of the information in speech is carried in the formant regions.

Although, as indicated above, all the information that is necessary for intelligible speech is contained the dynamic spectral plot, it is difficult to extract this information by visually examining the spectrum. An exception to this would be the location of the formant regions which appear as peaks. A sound spectrogram, however, allows us not only to observe the formant regions but practically all of the other interesting aspects of speech. The sound spectrogram is a plot of three variables; frequency, time, and intensity. The third dimension is indicated by shading; the heavier the shading the higher the intensity. Figure 2 is a typical Spectrogram reproduced from "Speech and Hearing in Communication" By Fletcher.

The spectrogram shows the formant regions of high intensity. It also shows the unvoiced sounds as a broad smear in the high frequency region. If the reader will say aloud the sentence beneath the spectrofram he will observe that the shifting resonance indicated on the graph

[7]E. W. Pappenfus, Ibid., p. 338.



—Voiced Speech Versus Whispered Speech. Sound Spectrograms of the Words "Speech we may see" Using a Wide-Band Analyzing Filter (300 Cycles).

Figure 2. Sound Spectrogram of Speech

corresponds to the changing of the resonant frequency by movement of the jaw. Thus while the vocoder demonstrates the importance of the intensity and frequency to intelligibility, the spectrogram verifies our other observations as to the nature of speech.

At this point we are ready to conclude that our analysis of speech is indeed a correct one. However, there is one question which comes to mind that should be answered. That is, if it is the intensity and frequency that are important to intelligibility how does a trained person get all of the necessary information merely by reading the senders lips? Obviously he is deprived of the "vital" ingredients of frequency and intensity.

To answer this question we will make reference to a simple electrical analogue. Assume the frequency spectrum of a given one syllable word to be correctly represented by Fig. 3.

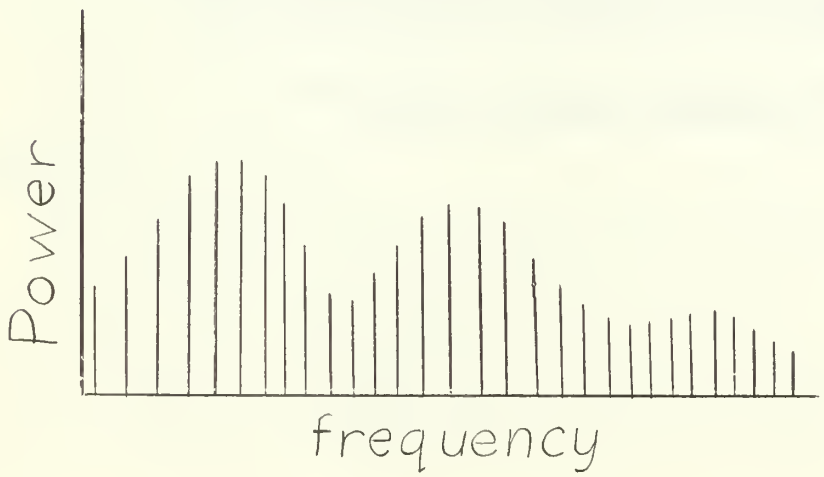


Figure 3. Frequency Spectrum of a Spoken Word

The formant regions are represented by the two peaks in the spectrum. As was previously indicated the shape of these formant regions is greatly affected by the manner in which words are started and stopped. The reason for this will soon become apparent. If desired, the shape could be defined as a function of the two dimensions, intensity and frequency. For now, assume that the most important of the two dimensions is the width (the assumption that the frequency is more important than the intensity will be justified in the discussion of hearing). Since the width of the formant region is frequency we will be justified in treating it as a bandwidth of a resonant electrical circuit. We know that the formant region is the important intelligence bearing region and since we have characterized this region by a single parameter, bandwidth, we may now think of the information as changing when the formant bandwidth changes. It is interesting to note that use of the bandwidth to represent the formant region is consistent with the theory that the absolute frequency values of the region are not as important as the relation between the beginning and ending (or bandwidth) frequencies of the formant regions.

By now it should be apparent how the bandwidth (i.e., the shape of the formant region) is varied. Recall that particular importance was given to the oral cavity because it could vary over wide limits. The oral cavity can be thought of as being analogous to a low Q resonant

cavity. As the shape of this cavity is changed by moving the jaw, tongue, and lips the Q of the cavity is varied. The jaw and tongue are most important in the formation of vowels while the lips play the important role in the formation of consonants. As was previously mentioned the consonants have a subtle effect on the shape of the formant region. Thus the shape of the power spectrum is directly related to the movements of the jaw, tongue, and lips. Therefore there is no flaw in the frequency-intensity concept of speech; the art of lipreading merely indicates that is possible to learn all of the physical movements associated with the various frequency-intensity patterns.

In conclusion, then, we can state that the necessary and sufficient characteristics of speech that must be retained for it to remain intelligible are its frequency and intensity information.

Hearing

In the preceeding discussion of speech, although not stated, it was tacitly assumed that the important properties needed for the recognition of speech were based to a large extent upon the requirements of the ultimate receiver, the human ear. Having determined from the investigation of the nature of speech that the frequency and intensity were the important factors it may now appear redundant or unnecessary to consider the hearing mechanism separately. This

would be true if we were only interested in linear speech processing operations. Clipping, however, being non-linear generates new frequencies and it also changes the amplitudes of the original frequency components. Thus, it is necessary to examine the theory of hearing to investigate such things as what effect undesired signals (such as those produced in clipping) have on the desired signal.

The ear consists of three sections, the outer ear (which collects the sound wave), the middle ear (which acts as a transformer between unlike media), and the inner ear. Only the inner ear will be of interest to our discussion. In its simplest form the inner ear can be considered to consist of the basilar membrane or corti, the liquid of the cochlea, the nerve endings (hair cells), and the nerve fibers.

The basic function of the inner ear is to receive the transformed speech wave from the middle ear and to convert it into nerve impulses so that the information can be transmitted to the brain. This is accomplished in the following manner: (1) vibrations of the basilar membrane (caused by the transformed speech wave) are transmitted through the liquid of the cochlea to the hair cells; (2) the vibrating hair cells generate an electrical potential (the so-called cochlear electrical potential) which sets off the nerve impluses; (3) these impulses then pass along the auditory nerve to the brain.[8] The auditory nerve consists of many

[8]Harvey Fletcher, op. cit., p. 110.

individual nerves and physically resembles a telephone cable containing many individual lines. The individual auditory nerves operate on an all or none basis, i.e. they are either on or off. When on, one nerve impluse is indistinguishable from any other. It has also been found that there exists a critical value of excitation necessary to change the nerve from one condition to the other. That is, there is a certain minimum value of excitation potential or intensity below which the nerve will not fire. Additionaly, it has been found that the amount of times a nerve can fire in a given instant is limited. Although the exact limit is not precisely known, it appears to be well below the upper frequency limit of audiability. The absolute frequency limit of each nerve is determined by its absolute refractory period (relaxation period between on-off condition). The absolute refractory period is the shortest relaxation period possible for the given nerve fiber. The preceeding facts about auditory nerves implies that the sensation of loudness is related to both the number of nerves firing and their rate of fire.

The description given earlier as to what takes place in the inner ear is generally regarded as the proper sequence of events. There is still, however, not complete agreement as to how each part of the sequence is accomplished. As might be expected there have been many approaches to the theory of hearing. As more factual information

about the human ear has become available most of these theories have become of interest only in the historical sense. The following discussion of hearing will be based largely on the space volley theory of hearing.[9] This theory is essentially a compromise between the two broad categories of the earlier theories.

The space portion of the theory is related to the manner in which the basilar membrane is divided according to frequency. That is, only certain portions of the membrane can support modes of a certain frequency. The further the distance from the origin (the membrane is connected at one end only) along the membrane the lower the frequency that can be supported. The basilar membrane can then be thought of as a series of frequency filters which convert the complex vibrations into their individual components. It is this membrane that performs the important task of transforming the sound pressure waveforms into their power spectral densities. Although the analogy of considering the basilar membrane as a bank of filters fits nicely into our idea of wave analysis and power speciral densities, it should be pointed out that these "equivalent filters" are not highly selective. In fact, even though only certain portions or spaces of the membrane are capable of supporting only particular frequencies, these spaces appear to be a broad band nature with some overlap with the other spaces. This may

[9]E. G. Wever, Theory of Hearing (New York: John Wiley & Sons, Inc., 1949), p. 189.

be the reason that the exact location as to frequency of the formant region was found not to be critical in regards to intelligibility.

The volley portion of the theory deals with frequency also, but in a different sense. The volley theory is an approach to the problem of the maximum frequency limit of the auditory nerves. In volley theory it is assumed that the relaxation times of individual nerves are sufficiently different so that although a patch of nerves may initially fire together that an instant later they are enough out of synchronism so that they are all firing at different times. Ideally then, if we consider 1000 firings per second to be the absolute maximum of each nerve, three such nerves operating out of synchronism could transmit the information of a 3000 cps frequency component. These patches of nerves are therefore capable of relaying exact high audio frequency information to the brain despite the limitation of the individual nerves.

Now with this prelimary information on the hearing mechanism completed let us consider some hearing phenomena which will be important to us in our study of intelligibility of clipped speech. Of primary importance is the effect of unwanted signals on the desired signal. This effect is usually described as interference or masking. It should be pointed out that some texts on hearing make a distinction

between interference and masking.[10] They describe interference as an effect between any two frequencies and masking as an effect that occurs only for frequencies that are close (about 500 cps). The phenomena of interference and masking are actually different; however, since the cause and effect is the same for both (i.e., the presense of strong unwanted signal alters or destroys the desired signal), the two phenomena will be considered the same throughout this discussion.

Consider now the simplified model of the inner ear shown in Figure 4.

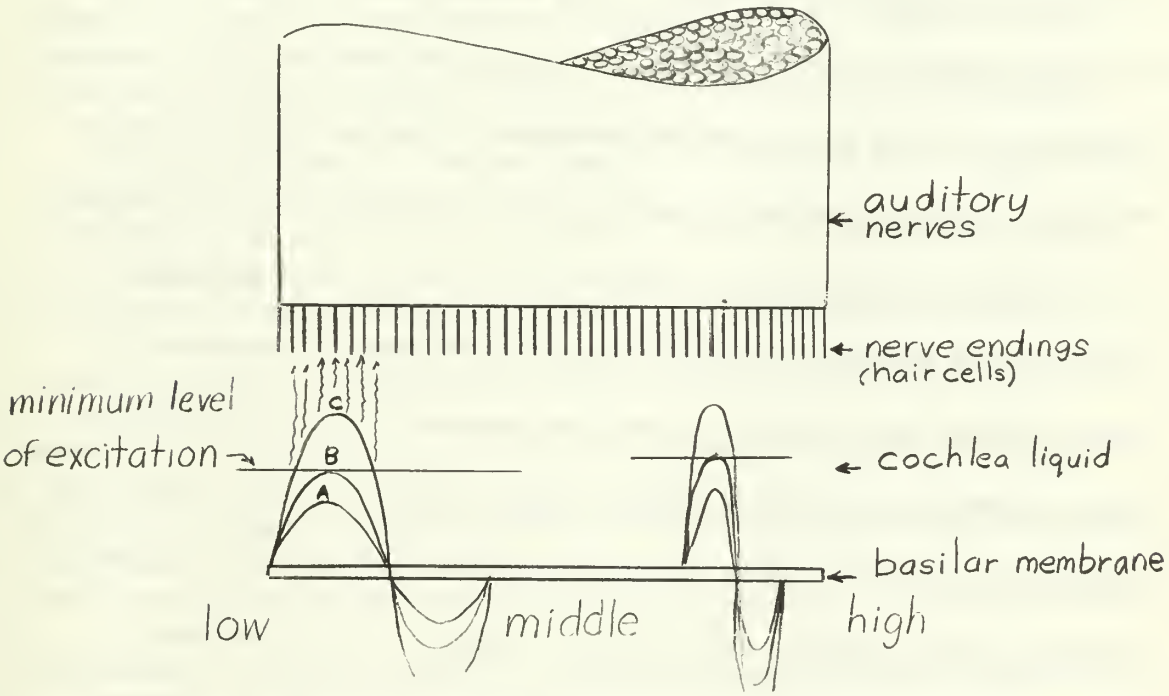


Figure 4. Model of the Inner Ear

[10]E. G. Wever, op. cit., p. 385.

Since the nerves will not fire until a certain minimum value is reached, the vibrations of the basilar membrane have to reach a certain minimum amplitude. Thus the low frequency sound which produced the vibration labeled 'A' in Figure 4. will not be preceived since it does not sufficiently disturb the hair cells to reach their minimum potential required to fire the nerves. The vibration labeled 'B is of sufficient amplitude to allow this tone to just become audiable. As the intensity is increased the amplitude increases and more and more hairs become affected by the vibrations. For the vibration labeled 'C', a large portion of the wave is above the minimum amplitude. Thus we can consider that all on the hair cells above this portion of the vibration are affected. This causes more nerves to fire and since the sensation of loudness of a sound is dependent upon the number of nerves firing and their rate, the loudness of the individual tone will increase with intensity. This statement should not be interperted to mean that the sensation of loudness is a linear relation with intensity since for complex waveforms which have many frequency components no such simple relation exists. About all that can be said of complex waveforms such as speech is the loudness is an increasing function of intensity. A look at a high frequency component will indicate why no simple relation exists for waveforms consisting of many different frequency components. Additionally, it should be emphasized that although the model is drawn to emphasize

the fact that more nerves are excited as the intensity is increased, there appears to be a region in which this increased nerve excitation causes no loss in intelligibility. This is verified from our own experiences since we know that it takes a great amount of amplification for us to notice any loss of intelligibility due to this increased power. This suggests that in this range the increase in nerve firings is essentially uniform with frequency; that is, the relative sizes of the patches of nerves that are firing appear to remain the same. This means we still receive the same message only the sensation of loudness is greater. Only when the intensities get into the very high range does the increased nerve excitation affect the intelligibility of the speech.

The high frequency waveform depicted in Figure 4. can be analysed in the same manner as was the vibrations in the low frequency region. The minimum level of excitation is drawn at a different level merely to indicate that the minimum vibration amplitude is not necessarily a constant for all of the nerves. It is instructive to note that the model correctly indicates that fewer hairs are influenced by the high frequency vibrations than for corresponding intensities for the low frequency wave. Additionally, there is a smaller increase in the percentage of hairs influenced as the intensity (amplitude) of the high frequency tone is increased. Since the vollwy principle tells us that more than one nerve is needed to transmit a high frequency tone, and since the number of

hair cells that can be excited by these high frequency tones are limited to a small area, an upper limit in the total number of nerve impluses per second is quickly reached as the intensity of the tone is increased. Further increases in the intensity will have negligible effect on the number of nerve impluses per second. This indicates why high intensity sounds are more easily preceived in the high frequency region than in the low frequency region. This is because the high intensity sounds in the low frequency region excite so many nerves that the original tonal relations become obscured. To better picture this consider the following. The basilar membrane acts to convert the complex waveform into its frequency spectral density. Now suppose that there are two formant regions; one in the low frequency region of the membrane, and the other in the medium frequency or mid-range. Because of their original high intensities (recall this is characteristic of formant regions) each will excite a patch of nerves. The rate of firings of the nerves relay the frequency information to the brain while in a rough sense the relative size of the patches of nerves firing indicates the intensities and therefore shape of the formant regions. As the intensity is increased the patch of nerves excited by the low frequency region will increase slightly faster than the mid-range patch of nerves. If the intensity is increased enough there will be reached a condition where the relation between the number of nerves firing in each of the corresponding

formant regions are so disportionate that the low frequency region becomes dominate and the original message is lost. It is obvious that this effect is less pronounced for the high frequency region where the percent increase in nerves firing due to increased intensity is at a minimum. This explains the fact that intelligibility increases as the frequency range and intensity level of a speech signal increases.[11]

Let us consider now how masking can occur. Suppose we have a desired signal at a given intensity. It will excite a certain number of nerve cells which will transmit a total number of impluses each second. Now if a second signal is introduced and if its intensity is continually increased, more and more nerves cells will be excited by this signal. It then becomes possible (in fact probable) for the total number of nerve impulses per second due to the second signal to become so large compared to the impluses/second due to the original signal that the original signal will no longer be perceived. This is masking. It is now easy to see why a low frequency tone (which is capable of excitation of large numbers of nerve endings) can easily mask a high frequency tone (which is capable of exciting only a small portion of nerve endings). Although not impossible, it takes unusual conditions for a high frequency tone to mask a low frequency tone.

[11]Irwin Pollack, "Effects of High Pass and Low Pass Filtering on the Intelligibility of Speech in Noise," Journal of Acoustical Society of America, Vol. 20, No. 3, 1948, p. 259.

In the discussion of speech it was stated that the consonants consisted of sounds which though random in nature had their intensities located in the high frequency portion of the spectrum. This coupled with the fact that the consonants are low intensity sounds means that consonants excite relatively few auditory nerves and are therefore easily masked by noise which is capable of exciting nerve endings anywhere on the entire auditory terminal due to its own random frequency spectrum. Thus even low intensity noise will excite many more nerve endings than most consonants and the probability is very high that in the presense of noise the consonants will not be preceived. By the same reasoning the vowels which are characterized by high intensity regions (formants) in the mid and low frequency ranges are not easily masked.

Masking can also occur when two tones of sufficiently different intensity have frequencies such that the vibrations that occur on the basilar membrane are very near. In this case the vibration of the less intense tone is essentialy lost in the larger vibration and the nerve endings receive no information about the less intense tone. When the tones are of nearly the same intensity the interaction between the vibrations on the membrane are such as to produce the phenomena known as beating. This represents an altering of the original information since a new tone is perceived in place of the two original tones.

We have only considered a small portion of the theory of speech and hearing and at that a greatly simplified approach. Still we should now have sufficient background to understand why clipped speech is intelligible and why such parameters as the signal to noise ratio play an important role in understanding intelligibility. This should aid us in designing more efficient electrial networks for the transmission of speech.

CHAPTER IV

THE EFFECT OF INTERMODULATION DISTORTION ON THE INTELLIGIBILITY OF CLIPPED SPEECH

It was concluded at the end of the second chapter that it was the IM distortion inherent in the clipping process that was the primary cause of the change in intelligibility as a function of clipping. Since most systems include noise, this necessarily includes the assumption that the signal to noise ratio remains constant. The preceeding discussion of speech and hearing indicated that the parameters that were necessary in recognizing different words or sounds were frequency and intensity. This supports the validity of the original conculsion since the effect of IM is to alter the intensities of the original frequencies and to introduce new frequencies of various intensities. It has already been shown that noise affects intelligibility in that it can mask the desired signals. Since the effect of noise on a signal is a simple concept it would be handy if IM products could be treated as noise. This is, in fact, how IM products are normally handled.[1] Since the fundamental characteristic of noise is its randomness, this assumption

[1]E. W. Pappenfus, op. cit., p. 328.

does require some justification. Particularily since there appears to be one case where the underlying assumption of randomness does not hold.

Speech itself is random in nature and therefore the frequency components of speech waveforms can be considered as random variables. The IM frequency components are functions of the original (or desired) frequency components and because functions of random variables are themselves random variables; we are justified in treating IM products as random variables. Just as in the case of noise, it is not necessary to establish the type of distribution that these random variables have since the central limit theorum, subject to minor constraints, allows us for large numbers of variables to assume that the limiting distribution of the summation of these components is gaussian regardless of the common distribution of the components. Thus, the assumption that IM products affect intelligibility in the same manner as noise is reasonable.

Let us now consider a special case where this random assumption is not correct. Since the majority of information in speech is contained in the formant regions let us consider for simplicity speech which consists of only vowels. If we allow the vowels to occur in a random fashion their corresponding formant regions will also be random. There is, however, one common property shared by all of

these formant regions. It is that each formant region can only occur in the region of a harmonic of the fundamental frequency of the voice. Generally this bit of information is not very useful in that the pitch of the voice is not usually known. However, consider what occurs when speech which is not bandlimited is clipped. We know that the fundamental frequency of speech will occur somewhere between 80 and 350 cps. This is in the range of high energy for normal speech. Thus, when we clip in this range we will clip the fundamental frequency and strong undersired signals will occur exactly in the formant region (recall that clipping produces the original frequencies, harmonics of the original, and sums and differences of these frequencies). Clipping the fundamental frequency insures us that a strong undesired signal will fall directly in the region of highest information content. This greatly increases the probability of interference with the original signal and consequently lowers the intelligibility. It is obvious that clipping in any region other than where the fundamental frequency of the voice is located will only produce random effects.

This suggests that if the low frequencies (i.e., 0-350 cps) are filtered out prior to clipping that this effect can be eliminated. This is, in fact, the case. Of particular interest is a study which conducted to determine the effect on intelligibility of speech in

noise due to sharp frequency limiting.[2] The results of this study show that the intelligibility of speech in noise is greater (67 per cent word articulation) when speech frequencies above 394 cps are passed prior to clipping than when the entire speech spectrum (0-9000 cps) is clipped (57 per cent word articulation). It should be noted that this particular effect of IM distortion on intelligibility only occurs for audio clipping. If the clipping is done at a higher frequency such as radio frequency this effect does not occur. The speech spectrum at rf frequencies is simply the audio spectrum translated to the higher frequencies; however, these translated frequencies are no longer harmonically related. Thus clipping the low end of the rf speech spectrum does not cause the undesired frequencies to coincide with the formant regions, in fact, these undersired products will fall outside of the pass band.

Having now considered the very special case in which the harmonic relationship of pitch of the voice with respect to both the formant regions and the clipping process causes the IM phenomenon to have non-random effects, let us now consider the general case of random effects. For this portion of the discussion we will assume that we have eliminated this special effect either by filtering or

[2]Irwin Pollack, "On the Effect of Frequency and Amplitude Distortion on the Intelligibility of Speech in Noise," Journal of the Acoustical Societs of America, Vol. 24, 1952, p. 538.

r-f clipping. We are then free to consider how IM products effect intelligibility when these products behave as noise.

Since intensity and frequency are the primary indicators of intelligibility let us begin by looking at the power spectrums of normal (unclipped) speech and of clipped speech. To help visualize this, consider first the power spectrum of a single sine wave. Its power spectrum is discrete; in fact, it consists of a single spike (or delta function) located at the frequency of the sine wave. For convenience we will assume throughout the discussion that the signal is across a one ohm resistor and therefore the magnitude of the spike will be one fourth of the amplitude squared. Now suppose we severely clip this sine wave signal. The resulting wave will be essentially a square wave and from fourier analysis we know that the spectrum of a square wave consists of a fundamental and an infinite number of harmonics. In the case of the clipped sine wave the fundamental will be equal to the original sine wave frequency. Thus even though many new frequencies are generated in the clipping process, the original component is still retained.

Let us now consider a slightly more complicated signal, the common two tone signal. This signal consists of the summation of two pure sinusoids of different frequencies and its power spectral density consists simply of two delta functions, one located at each

frequency component. The power spectral density of the clipped two tone signal will consist of spikes at the two original frequencies and all the IM frequencies. The magnitudes or intensities of the frequency components will depend on the characteristics of the clipper. However, it will be shown in the mathematical analysis of the clipping process that although the two original signal frequencies are retained their relative intensites are not necessarily the same. The IM products are, of course, always of less magnitude than the desired signal frequencies.

This analysis could be carried on for more complex signals such as three, four, and five tone signals. Finally as we considered signals with large numbers of different tones the signal would begin to resemble speech signals. Fortunately, it is not necessary to do this since we can see the trend from the first two simple examples. It is this: that the power spectral density of the clipped signal will always contain the frequencies of the original signal plus many new frequencies (the IM products). As has already been pointed out, however, the relative intensities of the original frequencies present in the clipped signal will not necessarily be the same as in the original signal. These slight differences in the individual intensities are not sufficient to cause any loss of intelligibility. This is because in the normal range of intensities the nerve patches excited by the frequencies

(excluding IM frequencies) in the clipped signal are essentially the same as those excited by the original signal. Of course, this will be true only as long as the intensities of the clipped signal do not fall below the minimum intensities required for nerve excitation. It is not very probable that this will ever happen since additional power is usually put into the spectrum of the clipped signal. That is, the clipped signal is amplified prior to transmitting. Thus we can generalize and state that the power spectral density of clipped speech is the same as the power spectral density of the original speech in the presense of noise (IM products).

This implies that we would not want to clip (considering intelligibility as the only criteria) in systems with high signal to noise ratios. This is because the clipping process generates its own noise (IM products) and in a low noise system this clipping noise will generally be greater than the system noise. If the predominant noise in the system is the noise due to the clipping then we cannot improve the signal to noise ratio. This is because the noise (IM products) is part of the signal (clipped speech) and any increase in the signal strength causes a corresponding increase in the level of the clipping noise level.

Conversely, if the signal to noise ratio is low, clipping is very desirable. In this case we can amplify the signal (clipped speech)

until the intensity of the noise due to the clipping (plus the original system noise) causes the amplified S/N ratio to be the same as the original S/N ratio.

Since for low S/N ratios the system noise is very high, we can increase the signal power by a large amount before the noise due to clipping becomes significant compared to the system noise. The intensities of these two separate random noise sources present in the system can be considered as cumlative. This is because in the S/N ratio in which we are interested, the noise power is an average noise power and not an instantaneous power. All of the above indicates two important facts: (1) the level of noise in a system prior to clipping is an important factor in the intelligibility of clipped speech, and (2) limiting factor in improving the intelligibility of clipped speech is the ratio of the power contained in the desired frequencies to the power contained in the IM frequencies.

The reason the IM products are the limiting factor in intelligibility of clipped speech is the following. We can always (assuming sufficient gain is available) amplify the transmitted signal to improve the S/N ratio when the noise is due only to the system. This increase in signal to noise ratio will increase the intelligibility of the transmitted signal. An example of what is meant by system noise would be the noise in the first stage of a receiver in a typical transmitter-

receiver communication link. It is easy to see from this example that system noise in independent of the signal. Thus increasing the signal will always improve the signal to noise ratio. However when the signal we are amplifying is clipped speech we amplify the IM frequencies (clipper noise) along with the original signal frequencies. Thus, even if the power in the clipped signal is so large as to make the power in the system noise negligible we are still limited by the signal (desired frequencies) to noise (IM frequencies) of the clipped spectrum. Thus in a noiseless system (or one with infinite gain available) the ratio of the power of the desired frequencies to the power of the IM frequencies is a good approximation of the expected intelligibility of the system. Since the strongest IM product generally dominates the weaker ones (due to masking) a first approximation to the expected intelligibility of a clipped system could be found from the ratio of the desired frequency component to the largest IM frequency component of a standard two tone test. However, since few systems approach the conditions of being noiseless or having infinite gain, the noise level of the system must be known before a reasonable estimation of signal intelligibility can be made.

We see then, that in general the effect of IM distortion on the intelligibility of speech is the same as the effect of noise on the intelligibility of speech. That is, the main effect of the IM distortion is

that of masking. We also noted that we must guard against that special case where the IM distortion causes destruction of the information in the formant region. Several methods were suggested for avoiding this problem. These first few chapters have been devoted to explaining some of the "whys?" of the problem of intelligibility of clipped speech. This was accomplished mostly by calling upon facts and theories which are in agreement with experimental observations. This is generally a lengthy, and in some cases not totally satisifying, procedure. However, it is a necessary one if the problems of voice communications are to be properly understood.

There is another much more direct, and for that reason probably more satisfying, method of approaching the problem of intelligibility of clipped speech. This is through the use of statistical mathematics. This approach is extremely important in the analysis of signals and a general outline of the mathematics involved, the assumptions, and the justifications is included in Appendix A. It should be pointed out that the mathematical model that comes from statistical analysis must lead to calculated results that are in agreement with experimental results. Thus the mathematical approach, although more acceptable to some, is not unlike the previous theories which derived their merit from agreement with experiment.

An article dealing exculsively with the statistical approach to the

intelligibility of clipped speech was published in 1954.[3] The main result of this paper was that it showed by use of cross-correlation functions that for certain random signals there was a marked similarity in the spectral content of the clipped and unclipped signal. Since the random signals considered were representative of speech signals this showed mathematical that the power spectral densities of clipped and unclipped speech are correlated (have a tendency to be alike) and this therefore implies that the intelligibility of the clipped signal should be very much like the unclipped signal. This agrees with our previous discussion which stated that the power spectral density of the clipped signal represented the original signal plus noise. That is, we would expect a high degree of correlation between the original signal and the same signal plus noise.

The amount of correlation between the two spectra can be expressed mathematical by a term called the coefficient of correlation. It was implied in the previously mentioned article that this term might be used as an indicator of the intelligibility of a system. However the same problem that was encountered in using the ratio of signal power to IM distortion power (actually this is essentially what the coefficient of correlation is a measure of) is also encount-

[3]J. M. C. Dukes, "The Effect of Severe Amplitude Limitation on Certain Types of Random Signal: A Clue to the Intelligibility of 'Infinitely' Clipped Speech," Proceedings of the IEE (London), Vol. 102-103, Pt. C, p. 88.

ered here. That is, nothing is given about the level of system noise. For a low noise system, the coefficient of correlation would probably give a reasonable estimate of the intelligibility. However in an enviorment with a high level of system noise this figure would be practically meaningless.

Perhaps a ratio that comes closer to a true indicator of intelligibility is the effective signal to "noise" ratio discussed in Single Sideband Principles and Circuits.[4] This ratio is give as $R = Sav/(N + D)$, where Sav is the average sideband power at the receiver, D is the effective noise power produced by the distortion products of the clipper, and N is the average noise power from all other sources. This ratio is exactly analogous to the earlier discussion presented in this paper. In fact, the authors make similar observations about the ratio defined above. That is, in the absence of noise at the receiver (N is approximately zero), clipping of the signal is undesirable because it introduces distortion products and reduces the intelligibility. Conversely, when the receiver noise is such as to virtually mask the signal, the clipping level should be increased to a point where the sum of the noise and effective distortion increases faster than the increase in average signal power. The most important observation, however, that the authors make about the ratio is

[4]E. W. Pappenfus, op. cit., p. 328.

this; "Since Sav and D are both functions of the clipping level, the intelligibility can be maximized by choosing the proper level of clipping for a given magnitude of noise power, N." This means that if maximum intelligibility is the criteria of the clipper system, there then exists an optimum clipping level corresponding to each noise level. In a separate study on SSB speech clipping, experimental results also indicated that an optimal clipping level existed for different noise levels.[5] The relation was empirical and of the form C(max) equals L plus 8db, where L was the signal to noise ratio (defined for a peak signal rather than rms) and C(max) was the maximum clipping level (i.e., further clipping at that noise level decreased the intelligibility). What is desirable is to take this optimization idea just one step further. That is, let us investigate the possibility of optimizing clipped speech when both the clipping level and the system noise is specified. This is perhaps a more realistic problem since it could easily turn out in the preceeding approach that the maximum clipping level would not be realizible (for example the clipping level may be limited by the average power limitation of the system amplifiers). The problem is then; how can the intelligibility be improved by only changing the manner in which the clipping

[5]Air Force Cambridge Research Laboratories, Speech-Signal Processing and Applications to Single Sideband (Bozeman: AD-276 850, Montana State College, 1962), p. 65.

is done? All of the preceeding discussions indicate that the way to accomplish this is to find some method to reduce the IM products that are generated in the clipping process. To do this a mathematical model of the clipping process will be extremely useful.

CHAPTER V

MATHEMATICAL MODEL OF THE CLIPPING PROCESS

The mathematical model of the clipping process may be formulated in several ways. It could be represented statistically in much the same approach as in Appendix A or it could be formulated by the transform method sometimes used for nonlinear devices. Which approach is chosen depends mostly on the characteristics of the clipping process in which we are most interested.

A very convenient method of studying the IM products generated in the clipping process is by expressing the process as a power series. Power series have been used extensively to describe the input-output characteristics of nonlinear devices. The power series approach has the disadvantage in that it often requires a large number of terms to accurately describe a given nonlinear device of process. This sometimes results in unwieldy manipulations when using power series to solve problems. The use of digital computers has, of course, helped to lessen this problem considerably.

Shown in Figure 5 are the input-output characteristics of both ideal and practical clippers. The ideal characteristic is "ideal" in

the sense that the function is linear up to some amplitude and from that point on it is limited to a constant value.

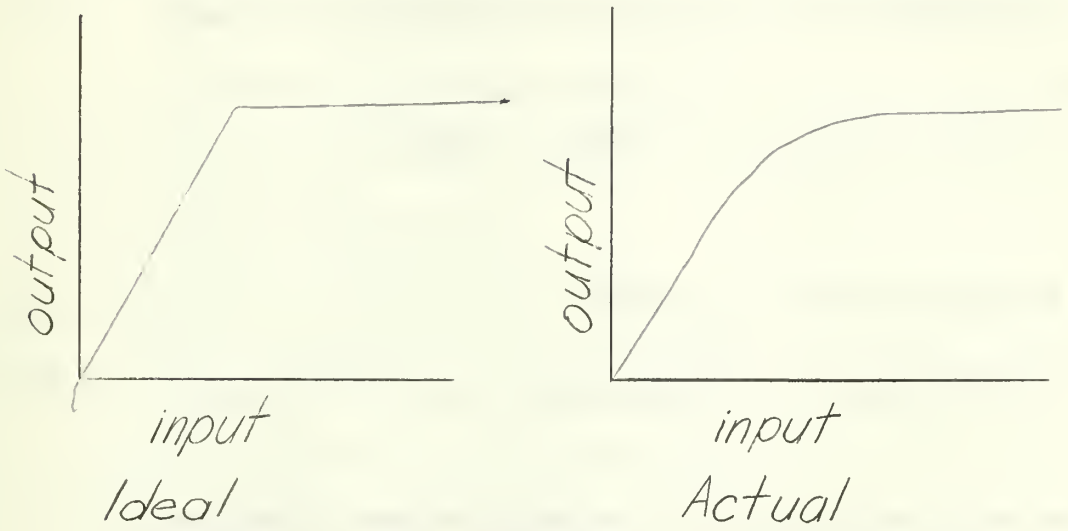


Figure 5. Clipper Characteristics

Accurate power series approximation over a large range of the clipper input-output characteristic is sometimes difficult to obtain. This is due to both the presence of the two relatively flat or linear regions and to the rather abrupt change between these regions. The problem of accurately describing the characteristic becomes more acute as the curves approach the ideal characteristic. In the limit (ideal characteristic) the point of discontinuity between the two linear regions make the power series approximation invalid. Despite these difficulties, reasonably good power series approximations can be made as the ideal limit is approached. That is to say that the actual

input-output characteristics encountered in practical devices (Figure 5) can be accurately approximated by power series representation.

Consider then that a clipper input-output characteristic is sufficiently described by the first five terms of a typical power series such as:

$$e_o = K_1 e_s + K_2 e_s^2 + K_3 e_s^3 + K_4 e_s^4 + K_5 e_s^5$$

If the input is a two-tone signal

$$e_s = A\cos\omega_1 t + B\cos\omega_2 t$$

The output, as described by the five-term power series, can be shown to be:[1]

d-c $(K_2/2)(A^2+B^2) + (K_4/8)(3A^4+12A^2B^2+3B^4)$

Fundamental
$$+\left[K_1 A + (3/4)K_3 A^3 + (3/2)K_3 AB^2 + (5/8)K_5 A^5 + (15/4)K_5 A^3B^2 + (15/8)K_5 AB^4\right]\cos\omega_1 t$$
$$+\left[K_1 B + (3/4)K_3 B^3 + (3/2)K_3 A^2 B + (5/8)K_5 B^5 + (15/4)K_5 A^2 B^3 + (15/8)K_5 A^4 B\right]\cos\omega_2 t$$

2nd order
$$\left[(1/2)K_2 A^2 + (1/2)K_4 A^4 + (3/2)K_4 A^2 B^2\right]\cos 2\omega_1 t$$
$$\left[(1/2)K_2 B^2 + (1/2)K_4 B^4 + (3/2)K_4 A^2 B^2\right]\cos 2\omega_2 t$$
$$\left[K_2 AB + (3/2)K_4 A^3 B + (3/2)K_4 AB^3\right]\cos(\omega_1 \pm \omega_2)t$$

3rd order
$$\left[(1/4)K_3 A^3 + (5/16)K_5 A^5 + (5/4)A^3 B^2\right]\cos 3\omega_1 t$$
$$\left[(1/4)K_3 B^3 + (5/16)K_5 B^5 + (5/4)A^2 B^3\right]\cos 3\omega_2 t$$
$$\left[(3/4)K_3 AB^2 + (5/4)K_5 AB^4 + (15/8)K_5 A^3 B^2\right]\cos(\omega_1 \pm 2\omega_2)t$$

[1] E. W. Pappenfus, op. cit., p. 181.

$[(3/4)K_3A^2B + (5/4)K_5A^4B + (15/8)K_5A^2B^3]\cos(2\omega_1 \pm \omega_2)t$

4th order

$[(1/8)K_4B^4]\cos 4\omega_2 t$

$[(1/8)K_4A^4]\cos 4\omega_1 t$

$[(1/2)K_4BA^3]\cos(3\omega_1 \pm \omega_2)t$

$[(3/4)K_4A^2B^2]\cos(2\omega_1 \pm 2\omega_2)t$

$[(1/2)K_4AB^3]\cos(\omega_1 \pm 3\omega_2)t$

5th order

$[(1/16)K_5A^5]\cos 5\omega_1 t$

$[(5/16)K_5A^4B]\cos(4\omega_1 \pm \omega_2)t$

$[(5/8)K_5A^3B^2]\cos(3\omega_1 \pm 2\omega_2)t$

$[(5/8)K_5A^2B^3]\cos(2\omega_1 \pm 3\omega_2)t$

$[(5/16)K_5AB^4]\cos(\omega_1 \pm 4\omega_2)t$

$[(1/16)K_5B^5]\cos\omega_2 t$

The frequency spectrum for two radio-frequency tones is shown below in Figure 6. This spectrum shows the frequency relationship for all of the components listed in the previous output equations.

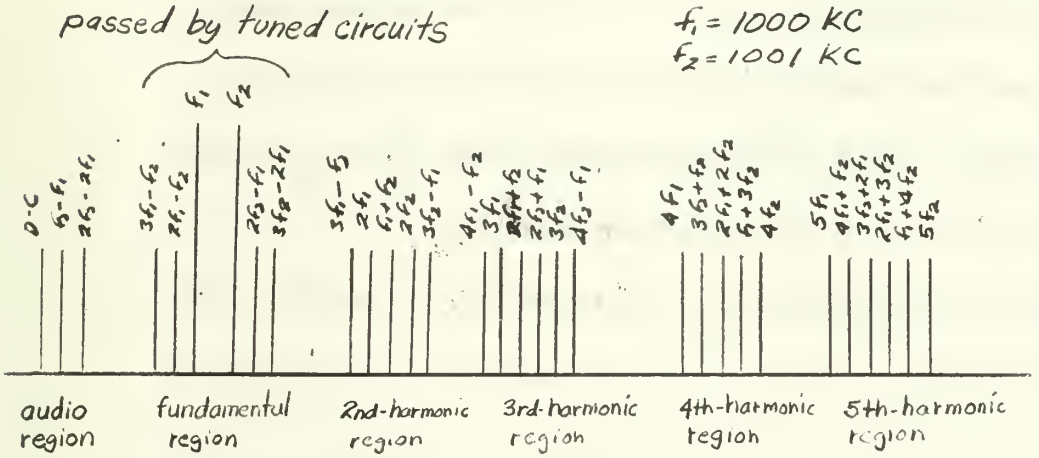


Figure 6. Frequency Spectrum Showing Relationship of Output Components.

The spectrum is representative of outputs resulting from R-F clipping. In this type of clipping only the odd order terms create distortion products which fall near the desired signal. The other regions are classified as harmonic regions and are far removed from the desired tones. For this reason a power series containing only odd order components is sometimes used to approximate the R-F clipping process.[2] However, all of the terms in the series must be used to represent audio clipping.

In audio clipping the same equations apply, therefore the same frequency components as shown in Figure 6 will occur. The difference is that the harmonic regions are no longer far removed from the desired pass band. For example if the audio pass band is considered to be 0 to 5000 cps, and the two audio tones to be clipped are 300 and 500 cps, then harmonics up to the 16th for the 300 cps tone and the 10th for the 500 cps tone will fall in the pass band. Associated with each harmonic region will be the distortion terms shown in Figure 6. Thus with audio clipping both the even and odd order terms cause distortion in the pass band.

In addition to indicating the frequencies that are present in the output of the clipper, the equations also indicate the magnitudes of

[2]Applied Research Lab. University of Arizona, Pre-Modulation and Post-Modulation Clipping in SSB Transmission (Tucson: AD-264 260, University of Arizona, 1960), p. 20; Air Force Cambridge Research Laboratories (Montana State College), op. cit., p. 89.

these frequency components. Examination of the equations for the two fundamental components indicates that when A = B (i.e., two tones of equal amplitudes), the two tones in the clipped spectrum will also be equal. For unequal tones the ratios will be slightly different in the clipped spectrum. This is dependent both on the magnitude of the coefficients and the ratio of the original two tones. In regards to magnitudes, the equations also indicate that in general the IM products due to the second order term are the largest. This is, however, highly dependent on the magnitude and sign of the coefficients. This last fact will be of importance when we consider optimizing the clipping process for maximum intelligibility.

CHAPTER VI

OPTIMIZING THE CLIPPING PROCESS FOR MAXIMUM INTELLIGIBILITY

Optimization

It has been stated that for a given signal to noise ratio there exists an optimum clipping level for obtaining the best intelligibility. That is, as we increase the clipping level the intelligibility increases until we reach a certain value (the optimum point). After this point, increasing the clipping causes a decrease in the intelligibility.

Now suppose that the system we are considering is physically limited as to the amount of conventional audio clipping that can be accomplished. Suppose further that when this limit on the clipping level is reached the intelligibility has not yet begun to fall off. That is, the optimum level of clipping has not been reached. What we would now like to know is can the intelligibility of the system be improved while maintaining a constant (the maximum of the system) clipping level?

The previous discussions indicate the most obvious method of accomplishing this is by clipping at higher frequencies (R-F). Clip-

ping at R-F instead of audio does not change the clipping level but it does remove many of the IM products. Let us then consider R-F clipping as a means of increasing the intelligibility of a system which is restricted as to both signal-to-noise ratio and allowable clipping levels.

Since we are considering R-F clipping as a technique for an audio speech processor rather than as a method of transmission, we have a choice as to the type of R-F signal we shall use. That is, should the R-F signal used be AM, DSB, or SSB? Single Side Band is, of course, the obvious choice since ideally the SSB signal represents the audio spectrum merely shifted up to the radio frequencies. This is precisely what is desired. The choice of any of the other modulation methods would be unsatisfactory because the signal that is to be clipped would be different from the audio speech signal (i.e., either the carrier, the other sideband, or both would be present in addition to the desired sideband).

Much interest has been generated in the problem of SSB speech clipping because of the peak to average problem encountered when conventional audio clipping is used in SSB equipment.[1] Theoretically

[1]Air Force Cambridge Research Laboratories (Montana State College), op. cit.; Applied Research Lab. University of Arizona, op. cit.; W. K. Squires and E. T. Clegg, "Speech Clipping for Single Sideband," QST, July 1964, p. 11.

a SSB system would need infinite amplitude capacity to pass an audio square wave (such as clipped speech). However in practical systems the bandwidth requirements keep the peak to average values finite. In fact, the peak to average ratios resulting from audio clipping in SSB band-limited systems is generally an improvement over the peak to average ratio of unprocessed speech (14.5 db PEP) for the same system.[2]

The results of the studies of SSB speech clipping indicate a definite increase in intelligibility as compared to audio clipping. This is in complete agreement with the previous discussion which stated that it is the presence of the IM products acting as noise that causes a decrease in intelligibility.

Of the studies mentioned, one of the most comprenensive of the group was a study conducted in 1962 by Montana State University at the request of the United States Air Force. Much of the results of their report are of interest and useful to the present inquiry. For example, the report gives results of tests comparing the intelligibility of speech in noise for a simulated SSB system when the speech is clipped at audio, DSB, and SSB. As expected, the SSB clipping yields the highest intelligibility. What may be unexpected to some is that the DSB clipping yields slightly lower intelligibility scores than the

[2]E. W. Pappenfus, op. cit., p. 318.

audio clipping. Except for mentioning that the IM or spurious products are noticeable more severe for the DSB clipping than for the SSB clipping, the report makes no attempt to explain the difference. It appears, however, that this reduction of intelligibility can be explained in terms of the IM products. Recall that it was stated earlier that it was not desirable to clip the DSB signal because it was a more complex signal than the translated audio signal (SSB). Let us consider the three types of clipping (audio, DSB, SSB) for a single audio tone. For the audio and SSB signal there will be no IM products, only the originial frequency and harmonics. The DSB signal however is actually a two tone signal and clipping it will cause IM products to fall in both the upper and lower side bands. In the case, of a two tone audio test signal, the DSB signal will actually be a four tone signal and there will be many IM products produced by clipping this signal. Thus clipping the DSB signal is not simply a matter of getting twice as many IM products as a clipped SSB signal. In fact, so many IM products are formed by the clipped DSB signal that despite the benefits of clipping at R-F, a sufficient amount of IM products occur in the passband to cause the clipped DSB signal to be less intelligible than the clipped audio signal. Thus the results of the DSB and Audio clipping are further experimental evidence that it is the IM products that cause loss of intelligibility.

It should be noted that the report indicated that all three types of

clipping gave improvement over unprocessed speech for the signal to noise ratios investigated. Typical levels of intelligibility for a signal to noise ratio of 15db (defined for a peak value of signal) and clipping level of 36db are; audio (87% articulation), DSB (81%), and SSB (94%).

Since the case of clipping a SSB speech signal instead of audio speech, as presented by the various studies previously mentioned, appears well established, we now ask now can this technique be implemented in an audio speech processing scheme. Such a system could be used with any type of communication mode including public address systems. The block diagram shown in Figure 7 represents a system which is an audio speech processor that uses SSB clipping techniques.

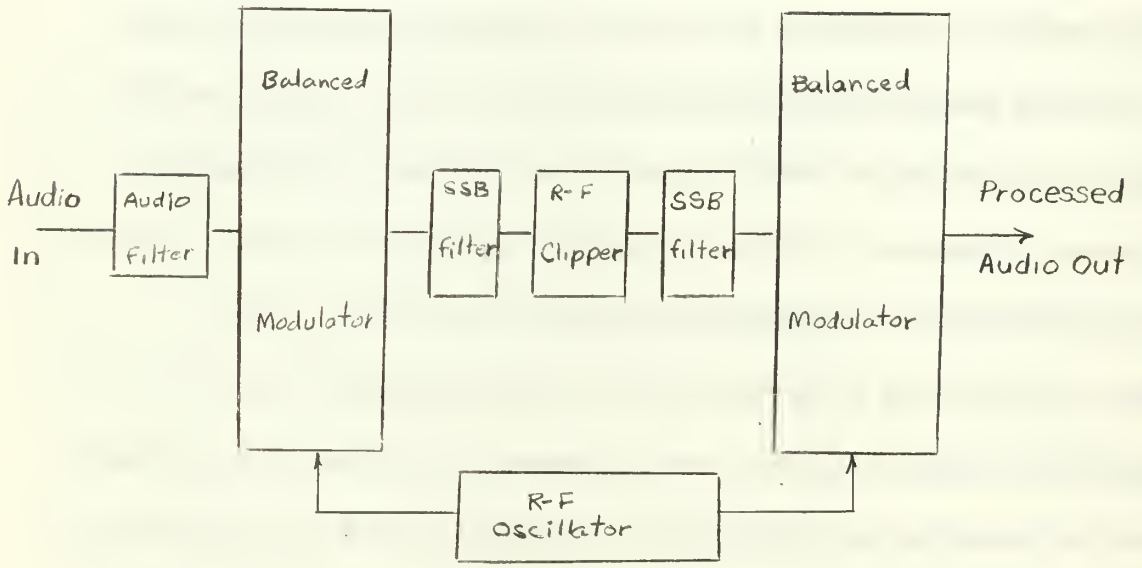


Figure 7. Audio Speech Processor Using SSB Techniques.

The scheme can be easily deduced from the diagram. The idea is to translate the speech signal to R-F by means of the first balanced modulator and the R-F oscillator. The filter converts the DSB signal into SSB; it is then clipped and brought back down to audio frequencies by the second balanced modulator. Thus this audio-to-processed-audio system could be used in many applications. There are some draw backs to this type of processor. The most obvious is that the system requires fairly precise components. For example precision filters and a stable R-F source are required to produce a good SSB signal, also the balanced modulator used would have to be designed, built and then adjusted to have very low intermodulation distortion. Although increased performance is not generally compatable with simplicity it seems beneficial to investigate the possibility of a more simple processor. This further investigation is stimulated by some interesting observations that were made while considering the possibility of using an R-F limiter in place of a convential clipper in the SSB speech processor. The effect on the IM products due to the difference of the input-output characteristics of the limiter and the convential clipper were considered.

Let us return to the mathematical equations (Chapter 5) that were generated to represent the IM products due to a two tone signal. The problem we are considering is still the same, that is, reduce the IM

products by any feasible scheme. From the equations it is obvious that if we reduce the coefficients in the power series we will reduce the IM components. It is equally obvious that we do not wish to reduce the odd order coefficients any great amount because they directly effect the magnitude of the desired or frequencies. Fortunately, however, the frequencies resulting from the even order powers depend only on the even order coefficients. This means that if it is possible to obtain a clipper input-output characteristic that is represented by a power series approximation which has some of the coefficients of the even order terms with unlike signs then the IM products can be reduced without reducing the desired frequency components! What must be determined now is whether or not the input-output characteristics of practical clippers can be satisfactorily represented by this particular type of power series.

Power series approximations of clipper characteristics generally have alternating signs for the coefficients of the series. This means that all of the even and all of the odd power terms will have the same signs. This is due to the saturation property of these characteristic curves. That is, if the second order term is positive (it, and in fact, usually the first order or linear term are generally positive) the third order term will be negative and the signs will continue to alternate. Thus the higher order terms will be substracted from the

lower (i.e., 3rd from 2nd, 5th from 4th, etc.) and the curve will tend to bend over (saturate) rather than rise. It seems reasonable to suspect, however, that if the bending over or saturation is gradual, the signs of the coefficients of the power series will not necessarily alternate. That is, the compensation due to each successive power term is not as critical as in the sharp saturation case. What we will do then, is to look at a group of input-output characteristic curves that vary from near ideal to the almost linear case such as shown in Figure 8 and study the magnitudes and signs of their power series coefficients.

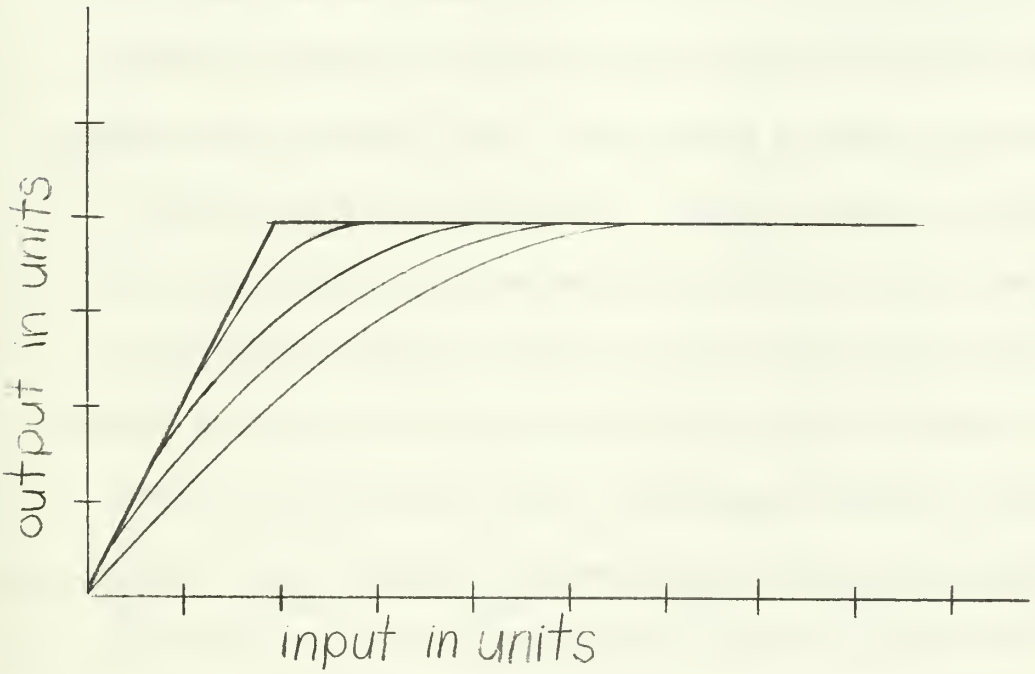
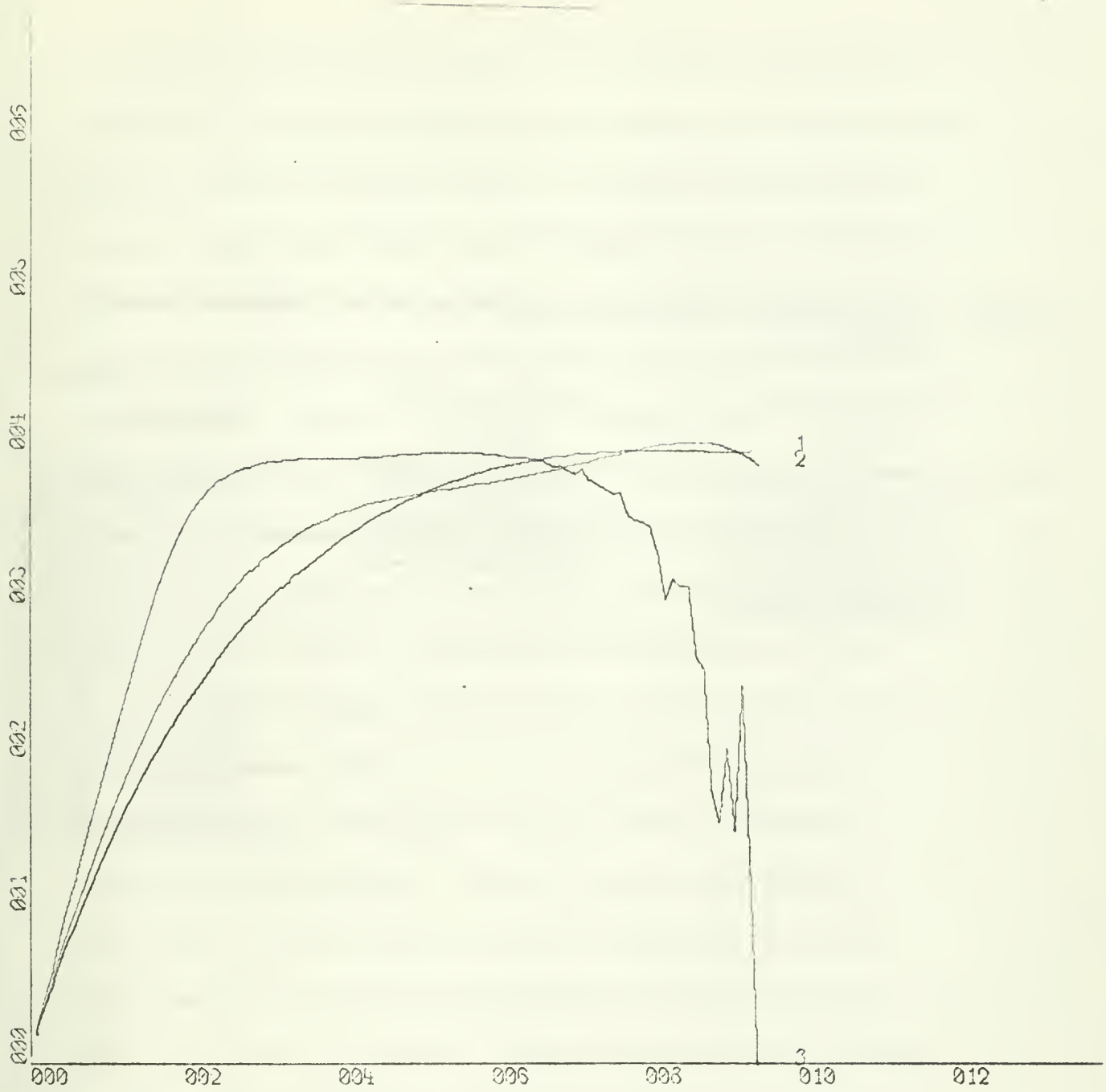


Figure 8. Clipper Input-output Characteristic Curves

To accomplish this the curves, such as those shown in Figure 8, will be used as data points and the power series approximation will be generated by a curve fitting computer program.[3] Then, by having the computer plot the calculated power series, the coefficients generated by the computer program can be compared directly to their corresponding curves. It should be noted that for comparison purposes all of the clipping characteristic curves which serve as the original data are kept at exactly the same clipping level of four units. This means that for large inputs the output is always constant (a straight line). This places severe demands on the power series approximation approach and on the computer program which is designed to convert the curves into matching power series. This is because in this region the curve is no longer nonlinear. The difficulty is due to the fact that in this region we are trying to approximate a straight line by a power series; this would require an infinite number of terms in the series. Therefore in this region the calculated curves have a tendency to oscillate or become degenerate. For the curves given, however, this perfectly linear region is not reached until about 6 or 7 units of input. Therefore for clipping levels below six units the calculated curves are sufficiently accurate representations of the original curves

[3]NPGS Computer Library, Program "Least Squares Curve Fitting with Orthogonal Polgnamials."



X-SCALE = 2.00E+00 UNITS/INCH.
Y-SCALE = 1.00E+00 UNITS/INCH.

LANNES SPEECH PROCESSING PROJECT

Figure 9. Plot of Three Calculated Input-Output Clipper Charateristics.

TABLE I. Coefficients of Power Series Approximations of Calculated Curves of Figure 9.

	COEFFICIENTS CURVE 1	COEFFICIENTS CURVE 2	COEFFICIENTS CURVE 3
CONSTANT	.3174715771E-01	.2153043929E-01	-.1343160607E+00
1	.1645268872E+01	.1738863788E+01	.4606624112E+01
2	-.3351950054E+00	-.1947598591E+00	-.1783464283E+02
3	.5182279024E-01	-.2432997753E-01	.6046790427E+02
4	-.6243238593E-02	.6315388720E-02	-.1189756476E+03
5	.5103356946E-03	-.3312896116E-03	.1489958244E+03
6	-.2595409676E-04		-.1267058990E+03
7	.6588831249E-06		.7656126391E+02
			-.3396196518E+02
			.1131514821E+02
			-.2873911870E+01
			.5610161605E+00
			-.8433957074E-01
			.9722162942E-02
			-.8493684884E-03
			.5504395455E-04
			-.2551646133E-05
			.7947516863E-07
			-.1476046131E-08
			.1217018692E-10

to study the desired characteristics. An actual plot of these calculated curves is shown in Figure 9 and the coefficients of their respective power series are shown in Table 1. It should be noted that the computer program used had a "best fit" feature. That is, when the addition of another term in the power series no longer reduced the error between the calculated curve and the original curve, the series was terminated. Notice that as the two extremes (ideal curve and linear) are approached it takes more terms to approximate the characteristics. Notice also as the ideal clipper characteristic is approached the coefficients become larger and their signs alternate (see curve 3). Curve 3 is typical of the characteristic of conventional clippers. The most important result, however, is depicted by the coefficients of curve 2. Curve 2, which is in the region of gradual curvature is seen to need only five terms to adequately describe its characteristics. Note that there are only two even power terms, the 2nd and the 4th, and that their signs are unlike. This is precisely the mathematical conditions to cause the even order IM products to be reduced. This power series approach to the reduction of the IM products is therefore mathematical feasible.

The important question is will physical devices having these characteristics behave in accordance with the mathematical prediction? That is, will a clipper that has a gradual slope to its input-

output characteristic produce less IM products than a convential clipper that has a sharp saturation characteristic (i.e., nearly ideal)? Before we can answer this, we must first find an actual clipping device that possesses the requisite input-output characteristic.

It might seem at first clance that we could always get the type of input-output characteristic desired by using piece-wise linear techniques. For our purposes, however, this does not seem like a profitable approach in that the piece-wise linear method introduces many points of discontinuities in forming the desired curve and it precisely these abrupt changes that we are trying to avoid. A more reasonable approach appears to be to use some device that already has the desired characteristic. With this in mind let us look first at the element used in the conventional audio clipper, the diode.

Recall that all semiconductor diodes require that a finite voltage be applied in the forward direction before they will fully conduct. The magnitude of this voltage is usually in the range of a few tenths of a volt. Thus, if we remove the external bias on a conventional solid-state clipper and if we keep the input signal small, the diode will function as a low level clipper. The advantage of this arrangement is that the zero-biased diode is not switched from the non-conducting to the conducting state as quickly as when an external bias is applied. In fact, the input-output characteristics

TABLE II. Results of Clipping Two Tone Signal with Conventional Clipper (Two Volt Bias).

IM COMPONENT (CPS)	500	1000	1500	2000	2500	3000	3500	4000	4500	5000
CLIPPING LEVEL (DB)	COMPONENT LEVELS IN DECIBELS									
4.6 DB	-23.9	-9.0	0.0	-29.2	0.0	-17.1	-23.2	-8.9	-32.2	-17.0
11.5 DB	-14.0	-10.2	0.0	-16.1	0.0	-22.6	-14.0	-10.0	-25.3	-32.2
15.9 DB	-11.4	-13.0	0.0	-16.5	0.0	-19.6	-11.4	-12.8	-18.4	-31.9
17.5 DB	-11.2	-12.6	0.0	-16.3	0.0	-31.2	-11.4	-12.6	-18.8	-27.0
20.4 DB	-10.6	-15.4	0.0	-18.1	0.0	-27.0	-10.5	-15.4	-16.2	-34.4
26.8 DB	-10.0	-19.5	0.0	-21.3	0.0	-23.0	-10.3	-19.5	-15.2	-27.5

TABLE III. Results of Clipping Two Tone Signal with Zero-Biased Clipper.

IM COMPONENT (CPS)	500	1000	1500	2000	2500	3000	3500	4000	4500	5000
CLIPPING LEVEL (DB)	COMPONENT LEVELS IN DECIBELS									
6.0 DB	-17.0	-45.0	0.0	-56.2	0.0	-56.2	-16.9	-45.0	-28.2	-55.8
8.6 DB	-14.6	-46.5	0.0	-59.5	0.0	-63.2	-14.3	-46.2	-25.9	-68.2
11.2 DB	-12.7	-63.5	0.0	-66.0	0.0	-60.8	-12.2	-49.0	-20.0	-67.0
17.5 DB	-11.4	-57.5	0.0	-57.2	0.0	-58.5	-11.1	-60.0	-16.8	-62.0
21.6 DB	-11.1	-63.5	0.0	-59.2	0.0	-59.2	-10.9	-60.0	-16.2	-61.5
27.9 DB	-10.6	-65.0	0.0	-66.2	0.0	-64.5	-10.4	-66.5	-15.5	-68.0
29.5 DB	-11.0	-65.0	0.0	-64.0	0.0	-62.5	-9.8	-70.0	-15.0	-67.0
34.8 DB	-10.6	-64.0	0.0	-67.5	0.0	-62.3	-10.2	-63.2	-14.9	-69.5
36.1 DB	-10.5	-64.0	0.0	-70.0	0.0	-72.0	-9.0	-72.0	-13.5	-72.0

of the zero-biased clipper is in close agreement with the gradual type of curvature required by the mathematical analysis. The results of clipping a two tone signal in the convential manner (diode with a two volt bias) is shown in Table II. Table III is the tabulation of the results of clipping a two tone signal with the same diode using no external bias (i.e., zero-bias).

The results shown in the two preceeding tables indicate that the zero-bias clipping method effectively removed the IM products due to the even order terms of the power series. For example, the sum and difference components, which are 2nd order effects, for 17.5db of conventional external biased clipping are both -12.6db. For 17.5db of zero-biased clipping the sum and difference components are -57.5db and -60.0db, which is so small that they are negliable. The slight discrepancy between 57.5 and 60.0 db is explained in Appendix B which deals with the equipment used and the method in which the two tone tests were conducted. The significiant reduction of the even order IM components by use of the zero-bias clipper is experimental evidence that the mathematical prediction is correct.

As was mentioned earlier the IM frequency components due to the second order term are generally the largest. Also recall from the discussion of the masking effect of tones that it is the strongest tones that predominate in the clipper noise. Thus, removing the

even order components will improve the intelligibility of speech by nearly the same amount as can be accomplished by R-F clipping which removes the low power harmonics in addition to the even order components. Note that neither zero-bias nor R-F clipping can remove the 3rd order components which are also strong.

Summary

This study has demonstrated that the intelligibility of clipped speech can be improved by clipping at low levels with no external bias on the diode clipper. This simple arrangement is very similar to the conventional audio clipping schemes, only the external bias is removed from the circuit and less amplification of the signal is required prior to clipping. We can expect that the zero-bias scheme will give improvement of the same magnitudes as was found for R-F clipping in the studies previously mentioned. This means that present systems utilizing conventional clipping can be further improved by adopting the zero-bias scheme. Additionally, new equipment can be designed with an even greater reduction of size and cost due to the decreased necessity of the equipment to handle peak powers. In the consideration of the design of new equipment it must be realized that accurate information as to how much improvement can be obtained from this scheme must be available. To accurately determine the improvement, extensive articulation or intelligibility testing

is required. It is therefore recommended that an extension of this present project be the undertaking of a series of intelligibility tests to determine if the magnitudes of improvement indicated in this report are attainable in practice.

In conclusion, it might be said that the main advantage of this approach to clipped speech is its simplicity. This simplicity manifests itself in two ways. First, this study allows us to approach the complex problem of the intelligibility of clipped speech with a single idea, the analysis of intelligibility as a function of the intermodulation products. Secondly, by considering the clipping problem as a simple power series approximation we can independently generate the same conclusions about the sharp and gradual clipper characteristics as are found by the more complex satistical mathematics approach.[4] Finally, these two thoughts combine to produce a practical solution to the problem, the zero-biased diode clipper.

[4]J. H. Van Vleck, and David Middleton, "The Spectrum of Clipped Noise," Proceedings of the IEEE, Vol. 54, No. 1, January, 1966, p. 2.

BIBILIOGRAPHY

1. Air Force Cambridge Research Laboratories. Speech-Signal Processing and Applications to Single Sideband. Bozeman: AD-276 850, Montana State College, 1962.

2. Cherry, Colin. On Human Communication. New York: John Wiley & Sons, Inc., 1957.

3. Davenport, W. B. and Root, W. L. Random Signals and Noise. New York: McGraw-Hill Book Company, 1958.

4. Dukes, J. M. C. "The Effect of Severe Amplitude Limitation on Certain Types of Random Signal: A Clue to the Intelligibility of 'Infinitely' Clipped Speech", Proceedings of the IEE(London), Vol. 102-103, Pt. C, 1955-56, p. 88.

5. Fletcher, Harvey. Speech and Hearing in Communications. New York: D. Van Nostrand Company, 1953.

6. Harmon, W. W. Principles of the Statistical Theory of Communication. New York: McGraw-Hill Book Company, 1963.

7. Licklider, J. C. R. and Pollack, Irwin. "Effects of Differentiation, Integration, and Infinite Peak Clipping upon the Intelligibility of Speech", Journal of the Acoustical Society of America, Vol. 20, No. 3, 1948, p. 42.

8. Office of Scientific Research and Development. The Effects of Amplitude Distortion upon the Intelligibility of Speech. OSRD Report No. 4217, Harvard University, 1944.

9. Pappenfus, E. W., Bruene, W. and Schoenike, E. Single Sideband Principles and Circuits. New York: McGraw-Hill Book Company, 1964.

10. Pickett, J. M. "Effects of Transmission Band and Message Contest on Speech Intelligibility", NRL Report 5690, U. S. N. R. L. Washington, D. C., 1961.

11. Pollack, Irwin. "Effects of High Pass and Low Pass Filtering on the Intelligibility of Speech in Noise", Journal of the Acoustical Society of America, Vol. 20, No. 3, 1948, p. 259.

12. Pollack, Irwin. "On the Effect of Frequency and Amplitude Distortion on the Intelligibility of Speech in Noise", Journal of the Acoustical Society of America, Vol. 24, 1952, p. 538.

13. Rasmussen, G. L. and Windle, W. F. (Editors). Neural Mechanisms of the Auditory and Vestibular Systems. New York: Charles C. Thomas Company, 1960.

14. Rosenblith, W. A. (Ed.). Sensory Communications. New York: Endicott House, 1961.

15. Squires, W. K. "The Computation of Single-Sideband Peak Power", Proceedings of the IRE, Vol. 48, Pt. 1, January, 1960, p. 123.

16. Squires, W. K. and Clegg, E. T. "Speech Clipping for Single Sideband", QST, July, 1964, p. 11.

17. Steinberg, J. C. "Effects of Distortion Upon the Recognition of Speech Sounds", Journal of the Acoustical Society of America, Vol. 1, 1929, p. 12.

18. Strassman, A. J. and Stockhoff, K. C. "Military Applications for Speech Compression Techniques", Hughes Aircraft Company, OP-24, April, 1960.

19. Tucker, D. G. "Intermodulation Distortion in Rectifier Modulators," Wireless Engineering, Vol. 31, 1954, p. 145

20. U. S. Army Signal Corps. Pre-Modulation and Post-Modulation Clipping in Single Sideband Transmission. DA 36-039-SC-80146, University of Arizona, 1960.

21. Van Vleck, J. H. and Middleton, David. "The Spectrum of Clipped Noise", Proceedings of the IEEE, Vol. 54, No. 1, January 1966, p. 2.

22. Wathen-Dunn, W. and Lipke, P. W. "On the Power Gained by Clipping Speech in the Audio Band", Journal of the Acoustical Society of America, Vol. 30, No. 1, January, 1958, p. 36.

23. Wever, E. G. Theory of Hearing. New York: John Wiley & Sons, Inc., 1949.

APPENDIX A

STATISTICAL ANALYSIS OF CLIPPED SPEECH

The ability of statistical mathematics to predict reasonable results in regards to the intelligibility of clipped speech depends primarily upon two principles of communications theory. The first is sometimes referred to as the Wiener-Khintchine relation, and the second is the principle of ergodicity.

The Wiener-Khintchine relation states that the spectral density of a signal is the Fourier transform of its autocorrelation function. The autocorrelation function is a measure of the likeness or correlation of a signal with that same signal displaced by time, τ. It is actually a time average of the product of these two signal (original and delayed signal) and is expressed as:

$$R(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t)x(t-\tau)\,dt \qquad (1)$$

The power spectrum or spectral density is defined as:

$$G(f) = \lim_{T \to \infty} \frac{1}{2T} |X(F)|^2 \qquad (2)$$

Where $X(f)$ is the fourier integral of the signal $x(t)$. In both equations, T is assumed to be an interval in which $x(t)$ is defined and finite but which is zero outside of this interval.

The fourier transform pair described by the Wiener-Khintchine relation is also valid for cross-correlation functions and cross-spectral densities. Cross-correlation functions are time averages of two different functions such as $x(t)$ and $y(t)$ displaced by a time difference . The cross-correlation transform relation is not as easy to derive as is the Weiner-Khintchine relation; however a brief outline as to how this relation may be derived is given by Dukes.[1]

The Fourier transform pair relation gives us a method of computing a signal's spectral density by evaluating time averages of the signal. For a certain class of signals this may also be done by what is called ensemble averaging. An ensemble (or ensemble of functions) is a set of random functions with a definite probability distribution defining the relative frequency of their occurance. The class of signals which are referred to here are ergodic signals. Ergodicity implies that the statistics of one system over a long

[1]J. M. C. Dukes, "The Effect of Severe Amplitude Limitation on Certain Types of Random Signal: A Clue to the Intelligibility of 'Infinitely' Clipped Speech", Proceedings of the IEE (London), Vol. 102-103, Pt. C, p. 88.

period of time are the same as the statistics of an ensemble of systems at one instant of time. Most signals that are of interest to us in communications are ergodic (for example, noise and speech). The cross-correlation function of ergodic signals may be expressed as an average of the product of random variables. Therefore:

$$R(\tau) = AV(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} XY\,p(X,Y)\,dX\,dy \quad (3)$$

Where the original signals $x(t)$ and $y(t)$ are now expressed as random variables X and Y, and P(X, Y) is the joint probability associated with the ensemble of random variables X and Y.

The advantage of this approach is that it is sometimes easier to correctly characterize the nature of a signal by a random variable and a probability distribution function than it is to characterize it in an interval of time. This is true of both normal speech and clipped speech. By representing both speech and clipped speech by appropriate random variables and by defining a joint probability function for these variables, the correlation function can be computed.[2] Then by using the principle of the Fourier transform pair, it is possible to formulate conclusions about the spectral densities of the signals. In the case of clipped and unclipped speech, Dukes com-

[2]J. M. C. Dukes, Ibid.

puted the cross-correlation function in the manner just described, and found the two signals to be highly correlated. This indicates that the power spectrum of the two signals (clipped and unclipped speech) are very much alike. Since intelligibility depends primarily on the power spectrum, the high degree of correlation is a mathematical indication that clipped speech should be very intelligible.

APPENDIX B

TWO TONE TESTS

A two tone test was chosen as the experimental method for determining the amount of intermodulation reduction that could be obtained by using a zero-biased diode instead of a conventionally biased clipper. A two tone signal is the most simple signal that will yield IM products when passed through a non-linear network.

The two tone generator used is shown in Figure A-1.

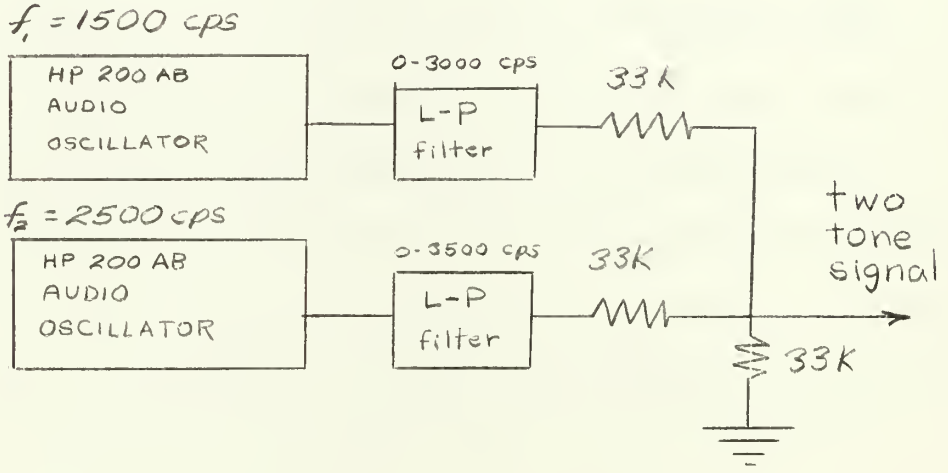


Figure A-1. Two Tone Generator.

The low pass filters are to aid in suppressing any harmonics generated in the audio oscillators. The 33K resistors serve as iso-

lation pads to reduce any direct mixing between the two oscillators. A simple ratio of 3 to 5 for the tones make the products easy to identify. With the arrangement shown above the following results were obtained.

	Frequency Compnent	Decibels
Test Tone	1500 cps	0
Test Tone	2500	0
	3000	-71
	3500	-72
	4500	-72
	5000	-72
	6500	-72
	7500	-62

With this two tone generator, two series of tests were run. The clipping circuit used was simply a two diode (IN337's by Raytheon) symmetrical clipper. The clipper had a two volt bias applied (symmetrically) for the first series of tests. For the second series of tests, the external biases were removed and the clipper functioned in the zero-bias mode. The results of these tests can be found in Table II and Table III of this paper. Ideally, the IM products formed on either side of the two tones should by symmetric in amplitudes. The results listed in the tables show some deviations from this ideal; these can be attributed in part to small drifts in the oscillator and wave analyzer circuits. A small difference could also be due to the fact that different tuned circuits are used in the wave analyzer for each frequency component measured. Examination of the results

will, however, show that these deviations are not significant.

The clipping levels indicated in the tables were obtained on a peak voltage basis. An example is shown below:

EXAMPLE CALCULATION

15 volts — peak to peak voltage of the two tone signal prior to clipping

4 volts — peak to peak voltage of two tone signal with clipper in circuit

Clipping level in db = 20log(15/4) = 11.5 db

The db levels of the IM products listed in the tables refer to the two tones (1.5 and 2.5 kc) and were recorded directly from the HP 302A wave analyzer.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	20
2.	Library U. S. Naval Postgraduate School, Monterey, California	2
3.	Commandant of the Marine Corps (Code AO4C) Headquarters, U. S. Marine Corps Washington, D. C. 20380	1
4.	Commandant of the Marine Corps (Code AO3C) Headquarters, U. S. Marine Corps Washington, D. C. 20380	1
5.	Prof Gerald D. Ewing (Thesis Advisor) Department of Electronics Engineering U. S. Naval Postgraduate School, Monterey California	5
6.	CAPT William Joseph Lannes, III, USMC 1045 Halsey Drive Monterey, California	1

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
U. S. Naval Postgraduate School Monterey, California	UNCLASSIFIED
	2b. GROUP Not Applicable

3. REPORT TITLE

INTERMODULATION DISTORTION: A CONTROLLABLE PARAMETER IN THE ANALYSIS OF THE INTELLIGIBILITY OF CLIPPED SPEECH

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Master's Thesis in Engineering Electronics

5. AUTHOR(S) (Last name, first name, initial)

LANNES, William J.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
May 1966	92	23

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Intermodulation Distortion Intelligibility Clipped Speech						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

DD FORM 1 JAN 64 1473 (BACK)